Integrative Methods

Integrative Genomics module

Michael Inouye
Centre for Systems Genomics
University of Melbourne, Australia

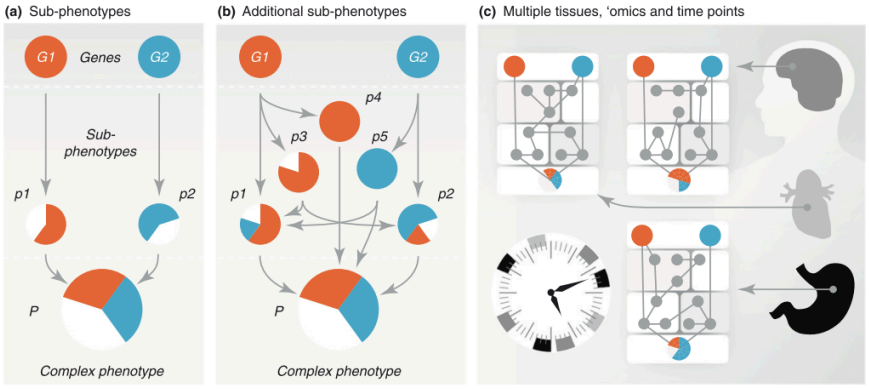
Summer Institute in Statistical Genetics 2016
Seattle, USA

@minouye271
inouyelab.org

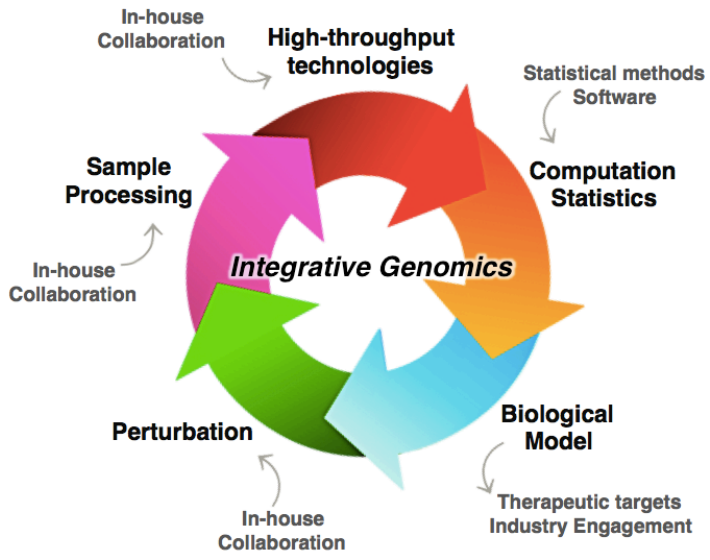
This lecture

- **Background**
- **Challenges**
- **Example: Gene co-expression networks in DILGOM cohort**
- **Example: Gene expression data to support genomics of metabolome**

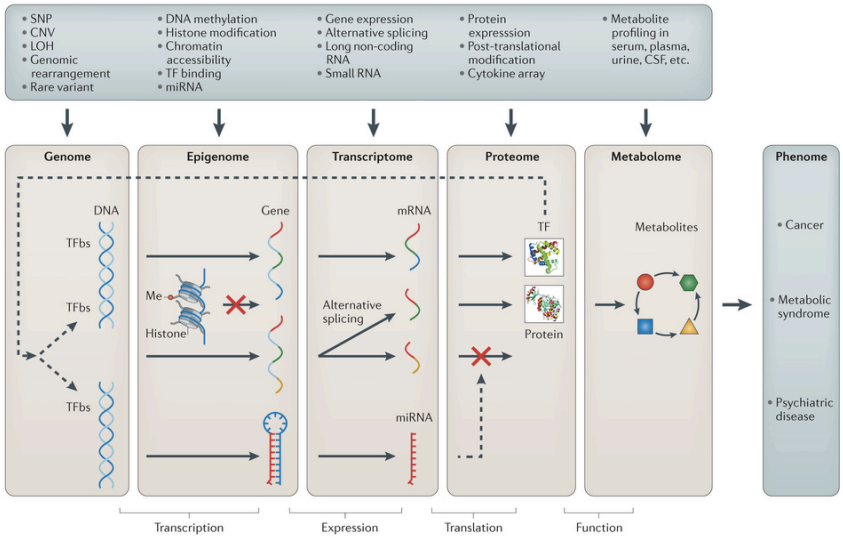
Background



Inouye, *Trends Genetics* 2011

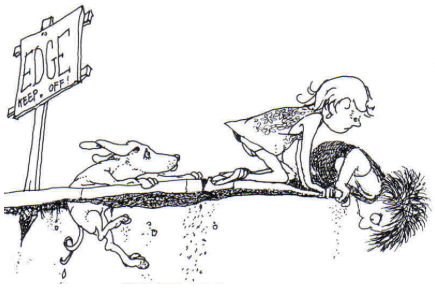


Multi-omics



Ritchie MD et al, *Nat Rev Genet* 2015

Where the sidewalk ends…



Integrative analysis is a relatively undeveloped area

Lots of scope for development and novel ideas

Nothing close to consensus on analytical approaches and strategies

Challenges

- **Large P: High dimensionality**
  - 10K, 100K, 10M variables per sample
- **Small N**
- **Heterogeneous data**
  - Different molecules
  - Different technologies
  - Different sampling strategies
- **Correlation**
- **Computational efficiency/feasibility**

Main things to be aware of

- **Understand the biological models underlying the data**
  - Context and interpretation
- **Know the technology**
  - Batches, biases, error profiles, sensitivities/specificities, missing data
- **Know the sampling strategy(s)**
  - Group-wise (case/control), population-based, enrichments, stimuli?
- **Spend time exploring the data**
  - Without exception, you will see things that require follow up
- **Build analysis pipelines and log all analyses**
- **The data may be complex but your analysis and presentation doesn't have to be**

Gene expression has a major role integrative analysis

- Insights into biomolecular networks
- Less technical variability than proteomics
- Relatively affordable
- Stable tissues are readily available
- Many network methods have been applied to gene expression data in the past
- Gene expression is thus a convenient way to characterise the average biological state of the cell population(s) being assessed

OPEN ACCESS Freely available online PLoS GENETICS

An Immune Response Network Associated with Blood Lipid Levels

Michael Inouye[1,2¤*], Kaisa Silander[3], Eija Hamalainen[1], Veikko Salomaa[4], Kennet Harald[4], Pekka Jousilahti[4], Satu Männistö[4], Johan G. Eriksson[4,5,6,7,8], Janna Saarela[3,9], Samuli Ripatti[3], Markus Perola[3], Gert-Jan B. van Ommen[2], Marja-Riitta Taskinen[10], Aarno Palotie[1,3,11,12], Emmanouil T. Dermitzakis[1,13], Leena Peltonen[1,3,11†]

518 randomly sampled individuals (DILGOM)

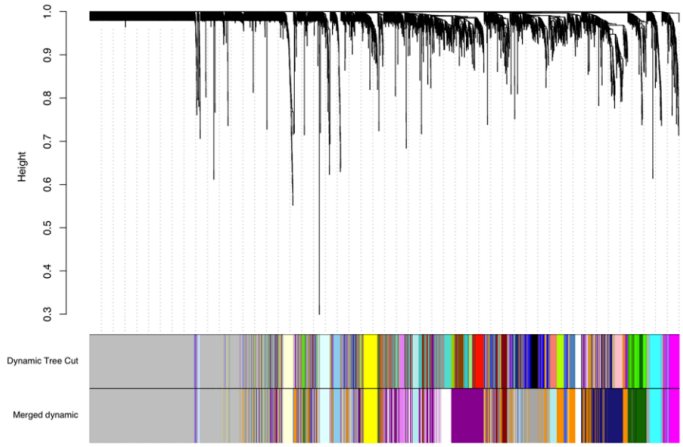
Fasting whole blood

Genome

Transcriptome

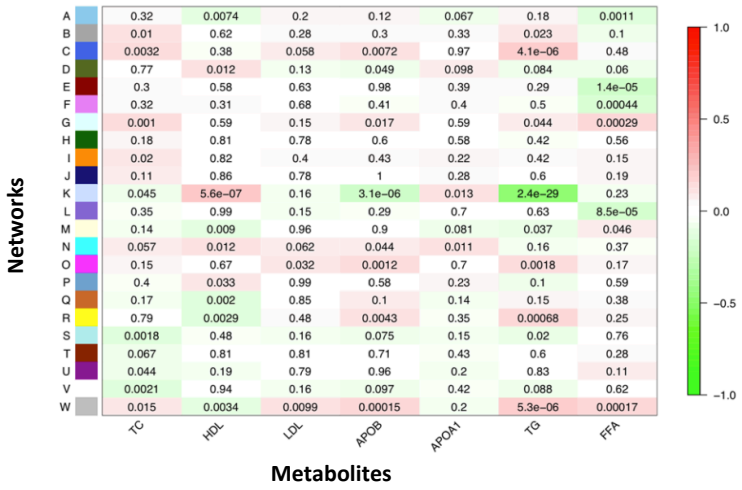
Metabolome

Identifying gene co-expression networks



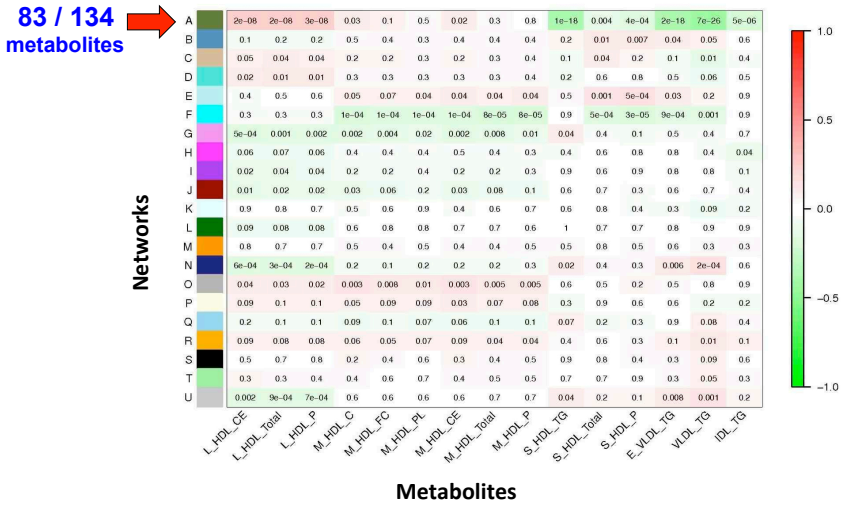
Inouye et al; *PLoS Genetics*, 2010

Networks and standard clinical lipid measures



Inouye et al; *PLoS Genetics*, 2010

Relationships between gene networks and metabolome



Inouye* & Kettunen* et al; *Molecular Systems Biology*, 2010

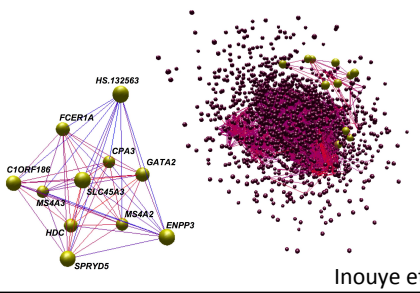
LL module appears to be involved in immune response

Genes

- ***FCER1A*** – high affinity IgE receptor
- ***MS4A2*** – high affinity IgE receptor
- ***HDC*** – enzyme for histamine synthesis
- ***CPA3*** – mast cell secreted peptidase
- ***GATA2*** – TF crucial for mast cell dev
- ***SLC45A3*** - *?*
- ***SPRYD5*** - *?*
- ***MS4A3*** - *?*
- ***ENPP3*** - *?*
- ***C1ORF186*** - *?*
- ***HS.132563*** - *?*

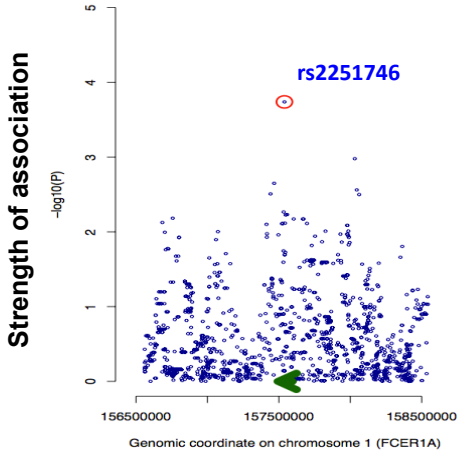
Immune markers

- IL-1ra (P=3.1x10^{-6})
- C-reactive protein (P=2.6x10^{-4})
- HMW adiponectin (P=1.6x10^{-5})
- Total IgE (P>0.05)



Inouye et al; *PLoS Genetics*, 2010

Does genetic variation influence LL module?



FCER1A $P = 1.83 \times 10^{-4}$
LL module $P = 4.28 \times 10^{-6}$

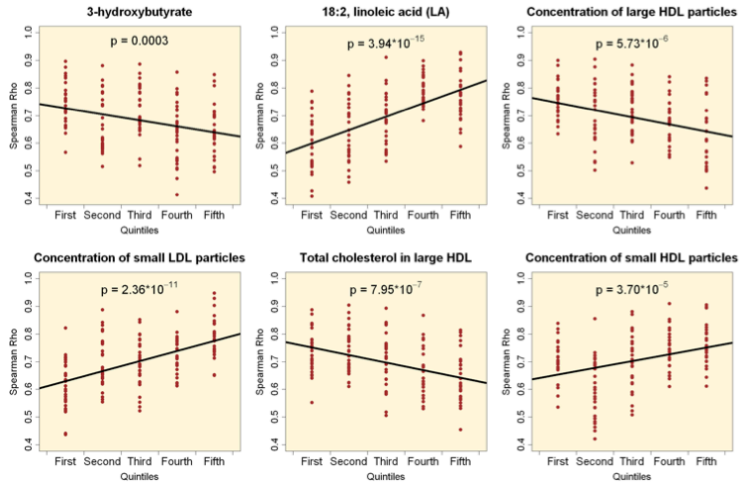
OPEN ACCESS Freely available online PLoS GENETICS

Genome-Wide Scan on Total Serum IgE Levels Identifies *FCER1A* as Novel Susceptibility Locus

Stephan Weidinger[1,2,3*], Christian Gieger[3,4*], Elke Rodriguez[2], Hansjörg Baurecht[2,5], Martin Mempel[1,2], Norman Klopp[3], Henning Gohlke[3], Stefan Wagenpfeil[3,6], Markus Ollert[1,7], Johannes Ring[1], Heidrun Behrendt[8], Joachim Heinrich[3], Natalija Novak[7], Thomas Bieber[7], Ursula Krämer[9], Dietrich Berdel[9], Andrea von Berg[9], Carl Peter Bauer[10], Olf Herbarth[11], Sibylle Koletzko[12], Holger Prokisch[13,14], Divya Mehta[13,14], Thomas Meitinger[13,14], Martin Depner[12], Erika von Mutius[12], Liming Liang[15], Miriam Moffatt[16], William Cookson[16], Michael Kabesch[12], H.-Erich Wichmann[3,4], Thomas Illig[3]

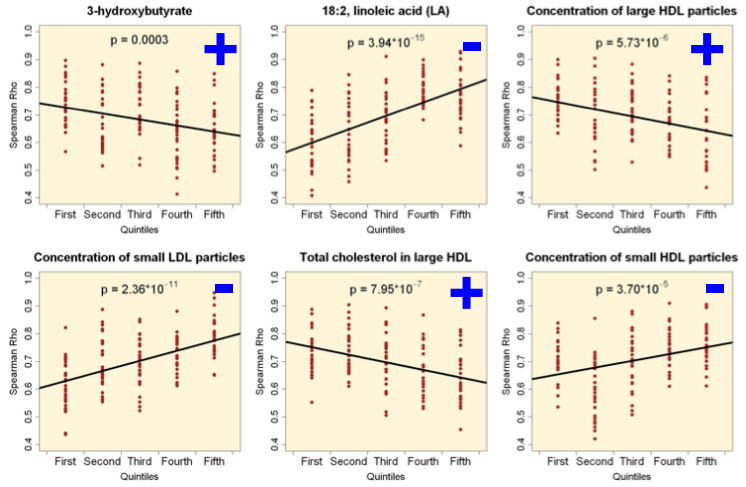
Inouye et al; *PLoS Genetics*, 2010

LL module appears reactive, do metabolites affect its connectivity?



Inouye* & Kettunen* et al; *Molecular Systems Biology*, 2010

Yes, it's a potential negative feedback loop



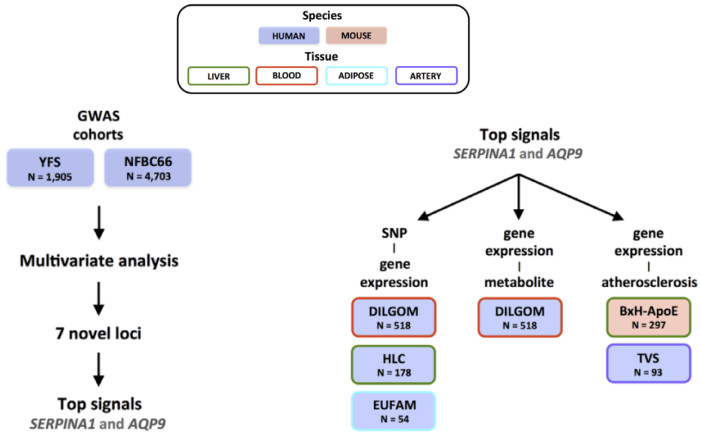
Inouye* & Kettunen* et al; *Molecular Systems Biology*, 2010

OPEN ACCESS Freely available online

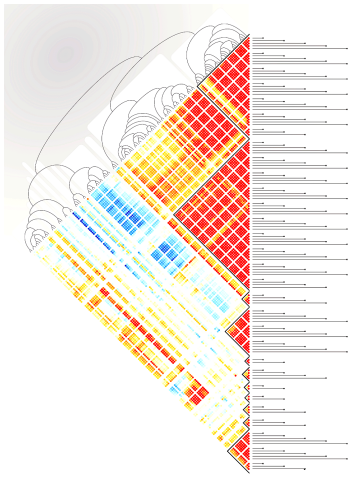
PLOS | GENETICS

Novel Loci for Metabolic Networks and Multi-Tissue Expression Studies Reveal Genes for Atherosclerosis

Michael Inouye[1,2*], Samuli Ripatti[3,4,5], Johannes Kettunen[3,4], Leo-Pekka Lyytikäinen[6], Niku Oksala[6,7], Pirkka-Pekka Laurila[3,4,8], Antti J. Kangas[9], Pasi Soininen[9,10], Markku J. Savolainen[9,11,12], Jorma Viikari[13], Mika Kähönen[14], Markus Perola[4], Veikko Salomaa[4], Olli Raitakari[15], Terho Lehtimäki[6], Marja-Riitta Taskinen[16], Marjo-Riitta Järvelin[11,17,18], Mika Ala-Korpela[9,10,12,17], Aarno Palotie[3,5,8,19], Paul I. W. de Bakker[19,20,21,22]



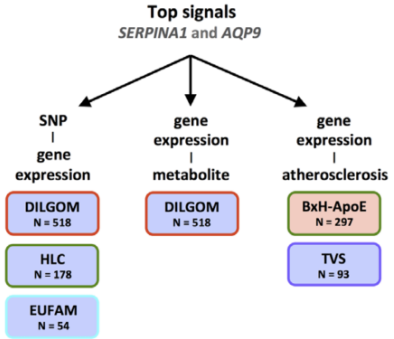
Determining metabolite networks



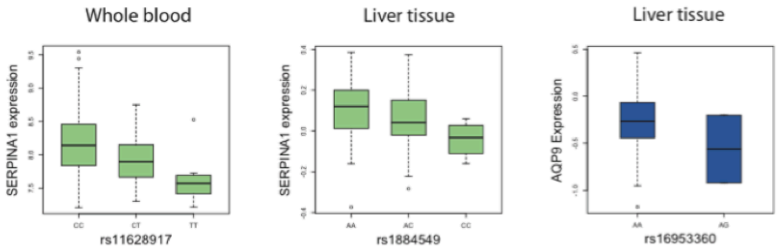
1 – apoB lipoproteins
2 – BC and aromatic amino acids large TG-rich VLDL
3 – large HDL
4 – small HDL
5 – polyunsaturated lipids
6 – ketone bodies
7 – glucose-alanine cycle
8 – renal function
9 – FA chain length/composition
10 – LDL diameter and FA composition
11 – urea & acetate

34 significant loci total 7 novel

Metabolite networks	Top SNP	Chr	Pos	Top Pvalue	Top metabolite	Gene
1,2	rs1303	14	93914596	5x10^-48	IDL-C	SERPINA1
1,2,3,4	rs16939881	15	56259271	3x10^-27	XL-HDL-TG	AQP9



SNPs for metabolic networks also drive *AQP9* and *SERPINA1* expression



DILGOM
N = 518
P < 10^-10
R^2 = 0.07

SNP associated with metabolic networks 1, 2

HLC
N = 178
P = 4x10^-3
R^2 = 0.04

SNP associated with metabolic networks 1

HLC
N = 178
P = 5x10^-3
R^2 = 0.04

SNP associated with metabolic networks 1,2,3,4

SERPINA1

Identification of genetic variants influencing the human plasma proteome

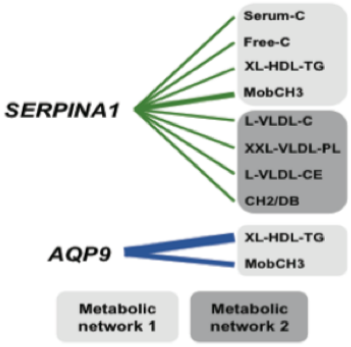
Åsa Johansson[a,b], Stefan Enroth[a], Magnus Palmblad[c], André M. Deelder[c], Jonas Bergquist[d], and Ulf Gyllensten[a,1]

[a]Department of Immunology, Genetics, and Pathology, Rudbeck Laboratory, SciLifeLab, Uppsala University, 75185 Uppsala, Sweden; [b]Uppsala Clinical Research Center, Uppsala University, 75237 Uppsala, Sweden; [c]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZC, Leiden, The Netherlands; and [d]Department of Chemistry–Biomedical Centre, Analytical Chemistry, SciLifeLab, Uppsala University, 75124 Uppsala, Sweden

Edited* by Richard N. Zare, Stanford University, Stanford, CA, and approved February 11, 2013 (received for review October 8, 2012)

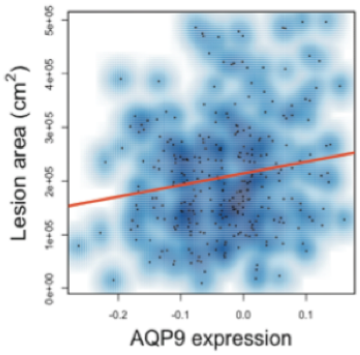
PNAS 2013

AQP9 and *SEPRINA1* expression is associated with metabolites



Metabolite	Chr	Position	Pvalue	Beta (95% CI)	Expressed Gene
XL-HDL-TG	15	56265176	8.48E-09	-0.61 (-0.82 - -0.41)	*AQP9*
MobCH3	15	56265176	7.16E-05	-0.43 (-0.63 - -0.22)	*AQP9*
MobCH3	14	93914570	6.46E-05	-0.51 (-0.75 - -0.26)	*SERPINA1*
L-VLDL-CE	14	93914570	2.47E-04	-0.49 (-0.76 - -0.23)	*SERPINA1*
XXL-VLDL-PL	14	93914570	2.48E-04	-0.51 (-0.78 - -0.24)	*SERPINA1*
L-VLDL-C	14	93914570	2.63E-04	-0.48 (-0.73 - -0.22)	*SERPINA1*
XL-HDL-TG	14	93924923	3.16E-04	-0.37 (-0.58 - -0.17)	*SERPINA1*
Free-C	14	93924923	3.98E-04	-0.37 (-0.57 - -0.16)	*SERPINA1*
Serum-C	14	93924923	4.00E-04	-0.37 (-0.57 - -0.17)	*SERPINA1*
CH2/DB	14	93924789	4.21E-04	1.14 (0.51 - 1.77)	*SERPINA1*

Liver *AQP9* associated with atherosclerosis in mouse model



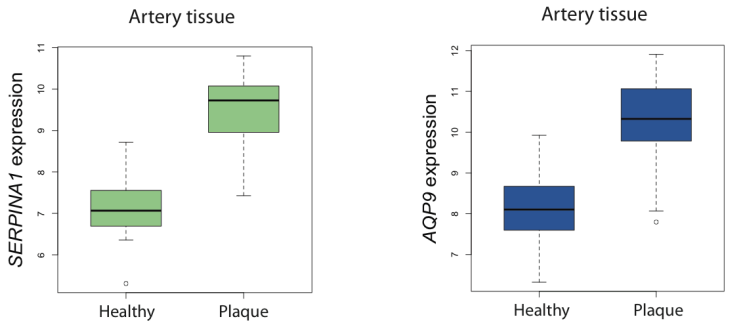
BxH-ApoE (N = 297):

- derived from backcross of highly susceptible to atherosclerosis (C57BL/6J *ApoE*$^{-/-}$) and highly resistant (C3H/HeJ *ApoE*$^{-/-}$).
- Fed on high-fat, western diet for 16 weeks then euthanized at 24 weeks.

***P* = 5x10^-3**

Samples in top decile of *AQP9* expression have on average 30% larger lesion area than those in bottom decile

AQP9 & *SERPINA1* in human aorta



Accessible resources for integrative genomics

- SageBase (via Sage BioNetworks)
- UK BioBank
- ImmGen
- ImmVar
- ENCODE
- THL Biobank
- TwinsUK
- iHMP / HMP2
- GTEx
- Epigenomics Roadmap Project
- Collaborative Cross (~outbred mice)
- Coming Soon: Precision Medicine Initiative